基于激光雷达数据的行人检测

任科飞，张 利

(清华大学 电子工程系，北京 100084)

摘 要：在自动驾驶领域涉及的众多任务中，行人识别是必须可少的技术之一。针对基于图像数据的行人检测算法无法获得行人深度的问题，提出了基于激光雷达数据的行人检测算法。该算法结合传统的基于激光雷达数据的运动目标识别算法和基于深度学习的点云识别算法，可以在不依赖图像数据的条件下感知和检测行人，进而获取行人的准确三维位置，辅助自动驾驶的控制系统作出合理决策。该算法在 KITTI 三维目标检测任务数据集上进行性能测试，在中等难度测试达到 33.37%的平均准确度，其表现领先于其他的基于激光雷达的算法，充分证明了该方法的有效性。

关键词：行人检测；激光雷达；点云；深度学习

中图分类号：TP391.41 doi: 10.19734/j.issn.1001-3695.2018.10.0786

Pedestrian detection based on LiDAR data

Ren Kefei, Zhang Li

(*Dept. of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)

Abstract: Pedestrian detection is a task which is necessary among all tasks leveraged in automatic driving domain. Traditional pedestrian detection algorithms took fully advantage of image data, unable to obtain depth of objects. To address aforementioned issue, this paper proposed a method based on raw LiDAR point cloud data. The proposed method combines traditional moving object detection in LiDAR data and point cloud recognition by deep learning, and is capable of perception and pedestrian detection without images, obtaining 3d location of pedestrian, therefore helping central control system make reasonable decisions. This method experimented in 3d object detection task of KITTI dataset, obtained 33.37% AP (Average Precision) on moderate cases. The results showed that proposed method performed better than other algorithm based on LiDAR data, which hence indicated the effectiveness of proposed method.

Key words: pedestrian detection; LiDAR; point cloud; deep learning

0 引言

行人检测需要解决的问题是在复杂的背景中定位行人，并用矩形框或立方体框标志其位置，因此行人检测属于目标检测的范畴。行人检测任务一直是计算机视觉领域中的热门问题和难点问题，在不同任务中都广泛应用，在自动驾驶领域中更是至关重要。自动驾驶技术需要通过传感器感知周围环境，并识别环境中的重要目标，如行人和车辆等，获取位置、深度、速度及运动轨迹等重要信息，辅助中央控制系统作出合理决策。深度信息表示目标到传感器的距离，是获取三维位置、速度等信息的基础。研究自动驾驶场景下的行人检测能够帮助系统作出更合理的决策，并避免交通事故和人员伤亡，因此重要性也不言而喻。

传统的行人检测算法主要利用图像数据，由于应用场景通常为室外，光照条件和天气情况会极大的影响图像质量，进而也会影响行人检测算法的精确度。Dalal 等人[1]提出基于 HOG(histogram of gradient)特征和支持向量机（support vector machine,SVM）的行人检测算法，利用 HOG 特征提供的边缘信息及梯度强度信息提供行人候选区域，结合 SVM 分类器进行分类。然而 SVM 分类器的计算复杂度非常大，与支持向量的数量成正比。Jones 等人[2]提出利用 AdaBoost 算法代替 SVM，提高了计算速度。RCNN 系列算法[3,4]通过卷积神经网络实现端到端多目标检测模型，在目标检测领域取得了非常优异的表现，然而，将 Faster-RCNN 模型直接用于行人检测任务效果并不令人满意，主要原因在于大多数场景下行人属于小目标，多层的卷积操作会造成目标的丢失，最终出现大量漏检情况。Zhang 等人[5]提出增加语义信息以及深度信息对 Faster-RCNN 进行改进，提高了 Faster-RCNN 模型的针对性，从而减少了漏检。基于图像的行人检测算法能够充分运用图像特征，在图像上定位行人的位置，然而无法获取行人的深度信息，虽然一些工作尝试通过深度神经网络估计目标的深度[6,7]，然而精度并不高。

随着激光雷达的广泛应用，可以在室外场景下精确的获取环境的深度信息。激光雷达通过激光脉冲对环境进行采样，获取环境到传感器的距离，受天气情况及光照条件的影响较小。激光雷达获取数据形式为点云数据，具有无序性、稀疏性以及特征少的特点，直接将在图像检测领域非常成功的卷积神经网络推广作为点云的处理工具并不能取得非常好的效果。Asvadi 等人[8]提出的运动目标检测算法将点云数据转换为体素，每个体素的值为 0 或 1，代表体素中是否含有点云数据，随后，通过在时间维度上的体素的累积判断目标是否为运动物体。Maturana 等人[9]提出了 VoxNet，将点云数据转换为体素形式，进而将卷积神经网络扩展为三维卷积神经网络对点云进行处理。然而由于点云的稀疏性，转换为体素形式后会出现大量零值，对识别任务毫无贡献，三维卷积也会对计算资源大量消耗。Qi 等人提出的 PointNet[10] 和

收稿日期：2018-10-18；修回日期：2018-11-28

作者简介：任科飞（1994-），男，河南淮阳人，硕士研究生，主要研究方向为图像处理、模式识别、激光雷达点云处理、点云分类（rkf16@mails.tsinghua.edu.cn）；张利（1965-），男，教授，博导，博士，主要研究方向为计算机视觉、模式识别.

PointNet++[11]是对点云分类算法的突破性创新。PointNet 首次直接对无序的点云数据进行处理，通过多层感知机提取点云全局特征进行分类；PointNet++是对 PointNet 工作的改进，增加了层级的点云区域特征提取子网络(set abstraction net)，提高了分类的精确率。然而，上述的点云分类只能应用于较小的场景，对于自动驾驶场景中大场景的点云数据并不能得到很好的结果。Qi 等人[12]提出利用图像二维检测器检测目标，将 PointNet 的应用范围限制在二维矩形框反投影得到的视锥内，级联 PointNet 网络进行二元点云分割，将点云分为目标和非目标点云，最后利用目标点云进行位置回归。文献[12]的缺点在于仍然依赖图像检测器，最后的检测结果受限于图像检测器的结果，也没有解决由于光照及恶劣天气带来的检测精度下降的问题。BirdNet[13]和 TopNet[14]为不依赖图像数据的车辆检测方法。这两种方法均通过对地面的估计，将点云映射到地面，将人为设计的信息转换为鸟瞰图数据，并利用 Faster-RCNN 网络对目标进行检测与识别。

本文提出的基于激光雷达的行人检测算法，能在不依赖图像数据的情况对点云数据进行处理，检测目标，获取目标的类别及位置信息。本文提出的算法拥有三个创新点：

a)提出了结合了传统激光雷达目标检测算法与点云分类算法的行人检测算法；

b)提出了密度自适应的地面估计方法，能够根据点云密度自适应地调整估计算法的应用范围；

c)提出了调整点云数据密度算法，对点云数据进行预处理，使其更适应传统的聚类算法。

1　地面去除

本文认为，需要检测的目标（行人、车辆、树木、建筑等）均位于地面之上，目标点云通过地面相互连接，因此移除路面点云后，目标点云将相互分离[15,16]，进而可以利用聚类算法即可获得行人候选聚类。传统的地面去除方法[17,18]将地面建模为一个固定的平面或者二次曲面，然而由于激光雷达本身的运动会对数据获取造成影响，通常情况下点云数据中的路面并不符合固定的曲面模型。文献[8]中将根据距离将点云数据分为多个部分并分别处理，构建不同的平面模型去除路面。然而，文献[8]中的方法采取固定的距离区间，当传感器的有效感知区间远小于设定的区间时（如建筑物遮挡时），在某些区间会出现全部是噪声或者不含有路面点云的情况，通过这些区间的点云估计地面是无意义的。因此本文提出了密度适应的地面去除算法。

1.1　密度自适应点云切片

激光雷达感知得到的点云数据具有密度不均匀的属性，且距离传感器的距离越远，点云数据的密度越低，且噪声越多。因此，将点云数据按照距离分为不同的部分，由于靠近传感器最近的部分数据密度最高，利用这一部分拟合的地面平面模型也最可靠。密度自适应点云切片的流程如下：

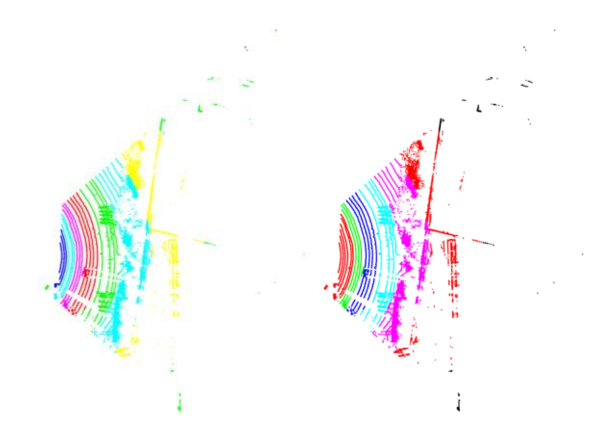
a）确定切片的最近距离与最远距离；

b）确定间隔参数λ，并确定切片间隔及个数 N；

c）判断切片中是否含有足够多的点云数据，即点云元素数量是否大于阈值 $Thresh$，如果小于 $Thresh$，则与下一个切面合并，重复此过程 N 次；

d）判断最后一个切片是否含有足够多的点云数据，如果元素数量小于 $Thresh$，则将此切片看做噪声丢弃。

图 1 展示的是文献[8]中的切片算法与本文提出的密度自适应切片算法结果的比较，图 1 中，(a)为文献[8]的结果；(b)为本文密度自适应切片算法的结果，不同的颜色表示属于不同切片的点云数据，黑色点表示被认为是噪声的点云。图 1(a)中最外层绿色的点云距离传感器最远，含有大量的噪声，不适合估计地面，然而在文献[8]中仍然会在后续过程中估计地面。在图 1(b)中，最外层点云由于点云元素数量小于阈值 $Thresh$，因此被当作噪声处理。可以看出，本文提出的密度自适应切片算法可以更有效的利用点云数据，提升了文献[8]切片算法的效率。



(a)文献[6]中的切片算法结果　(b)本文提出的密度自适应的切片结果

(a)slicing result of paper [6] (b)depicts proposed adaptive slicing result

图 1　两种切片结果比较

Fig. 1　Comparing results of two slicing approach

1.2　门限与拟合地面

对于每个切片中的点云数据，本文利用四分法确定门限。首先利用切片中点云高度的中位数 Q_2，将点云分为上下两部分；随后，分别确定上下两部分点云高度的中位数 Q_1 和 Q_3；最后，本文把高度大于 Q_3 且小于 Q_2 的点作为拟合地面的内点。这样做不仅能减少后续处理点云数量，减少处理时间，而且可以去除大部分噪声，提升算法效率及精度。然后，本文利用 RANSAC 方法[19]在每一个切片中通过内点拟合地面平面。RANSAC 是一种通过迭代寻找最优模型的拟合算法，对于含有较多噪声的数据，RANSAC 比直接最小二乘算法更加鲁棒。最后通过文献[8]提出的验证方法验证每个切片的地面模型。

2　聚类

聚类算法作为非监督学习算法已经被非常广泛的应用于数据挖掘，图像处理以及模式识别等领域。DBSCAN 算法[20]是一种鲁棒的基于密度的聚类算法，相对于其他聚类算法的优势在于 DBSCAN 不需要预先确定数据中的目标数目。DBSCAN 算法的聚类半径以及聚类中最小元素个数是全局唯一确定的，因此通常适用于密度均匀的数据，当应用于密度不均匀的数据，如激光雷达数据时，则聚类效果比较差。因此，笔者提出通过对点云数据的预处理，对其密度进行变换，使其在高度维度上密度均衡，再利用 DBSCAN 进行聚类处理，可以提高聚类的准确度。

2.1　预处理

激光雷达的扫描线通常以固定的垂直角度间隔分布，因此必然造成距离传感器近的数据密度高，而距离传感器远的数据密度低的情况。图 2 所示的为激光雷达的扫描线模型，图中 $z=0$ 表示激光雷达传感器所在的平面（通常安装在载具上），蓝色虚线为激光雷达发出的扫描线，d_1 为点云数据到传感器的最小距离，$\Delta\alpha$ 为相邻扫描线的间隔角度。假设在距离传感器 d_2 处存在一个点（图 2 中红色点），本文提出利用如下方法调整点云的高度维度。

首先根据相似三角形原理可以得到式(1)。

$$\begin{cases} \dfrac{h_2}{h_1} = \dfrac{d_2}{d_1} \\ h_2^{new} = h_1 \end{cases} \tag{1}$$

化简后可得到式(2)。

$$h_2^{new} = \frac{h_2}{d_2} \times d_1 \tag{2}$$

其中：h_2 为此点的原始高度坐标，h_2^{new} 为调整后的高度坐标。

图 3(c)为密度调整后的点云数据，可以看出，点云数据只有在高度维度上的密度放生了变化，这样不仅可以更适合聚类算法应用，而且能够保留点云数据中的几何特征。

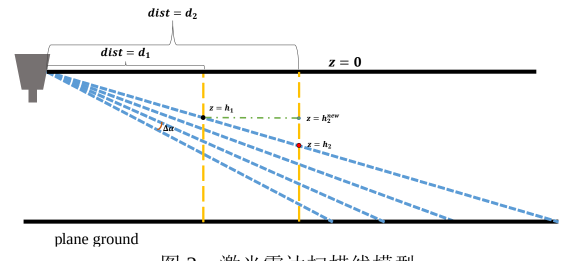
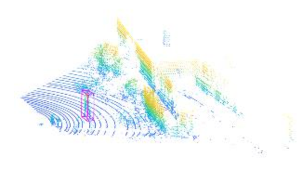


图 2 激光雷达扫描线模型

Fig. 2 LiDAR scan model

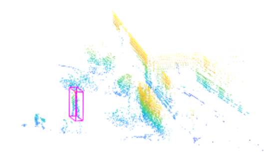
2.2 DBSCAN 聚类

DBSCAN 算法需要两个全局参数：eps 和 MinPts，分别表示聚类半径和聚类的最小对象数量。如果两点的距离大于 eps，则两点被认为不属于同一个聚类。如果聚类中的点数量小于 MinPts，那么此聚类被当作噪声。本文在实验中发现当 eps 设为 0.13，MinPts 设为 5 时可以得到最好的聚类效果。图 3 为点云数据在上述一系列操作中的结果。图中，(a)为原始的点云数据，红色边界框内为行人，点云的颜色根据点云高度坐标分配，蓝色为高度坐标较低的点，黄色为高度坐标较高的点；(b)为地面去除后的剩余点云数据；(c)为经过密度调整后的点云；(d)为 DBSCAN 聚类的结果，相邻聚类以不同颜色表示，黑色代表被认为是噪声的点，可以看出红色边界框内的行人显示为绿色，说明被当做行人候选聚类，另外大量背景也被认为是候选聚类，因此需要对候选聚类进行筛选。



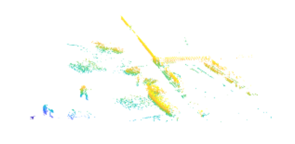
(a)原始点云数据(位于摄像机 FOV 内部分)

(a)raw point cloud data plane(only in FOV of camera)



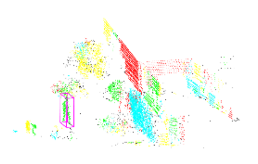
(b)去除地面后的剩余点云

(b)remaining point cloud after removing



(c)调整密度后的点云

(c)point cloud after density adapt process



(d)聚类操作得到的点云聚类

(d) represents cluster result

图 3 点云操作结果(不同颜色代表不同聚类,黑色为噪声)

Fig. 3 Results of a serial of point cloud processing(different color points belong to different clusters, and black points is deemed as noise)

2.3 筛选

根据行人的先验知识，可以对得到的点云聚类进行筛选以去除大部分背景聚类。这样不仅可以降低后续的计算复杂度，又可以去除大量的负样本，减少正负样本的不平衡程度。根据候选聚类的长度（l），宽度（w）及最大高度（h，候选聚类中高度最高的点距离地面的高度），本文通过以下规则筛选候选聚类，符合如下条件（式（3）~（6））的被认为是行人候选聚类，否则被认为是噪声。

$$0.01 < l < 1.0 \tag{3}$$

$$0.01 < w < 1.0 \tag{4}$$

$$0.5 < h < 2 \tag{5}$$

$$\sqrt{l^2 + w^2} > 0.1 \tag{6}$$

3 聚类识别网络

本文改进了文献[12]中的点云分割网络及回归网络作为识别网络。文献[12]中分割网络与回归网络级联，分割网络的输入为由二维检测结果限定的视锥内的一部分点云数据，输出为目标点云，回归网络的输入为目标点云，输出为边界框的参数。本文将分类网络和回归网络并联，输入均为行人候选聚类，利用 PointNet++网络作为特征提取网络，通过 set abstraction net 进行层级的局部特征提取，最后得到点云全局特征向量，接着利用全连接网络输出最后的结果向量。分类网络输出为 2 维向量，分别表示聚类属于背景和行人的概率，两者的和为 1，当属于行人的概率大于 0.6 时，将候选聚类判断为行人。回归网络会回归边界框的位置以及朝向，本文按照文献[12]的方法，将其分为中心坐标，旋转角度以及标定框尺度。旋转角度的回归分解为旋转区间的判断与残差回归。因此，回归网络的输出为 $(3+N+N+3)$ 维向量，其中 $(N+N)$ 维表示旋转区间的概率和残差。

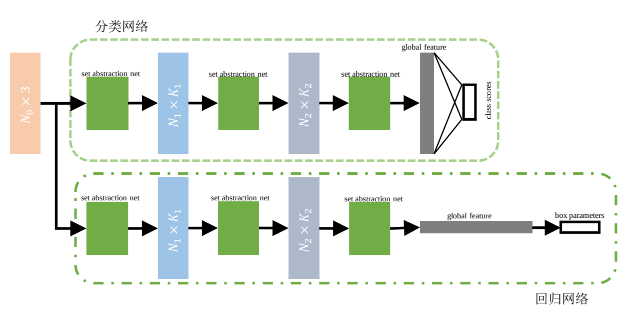


图 4 识别网络模型

Fig. 4 Point cloud recognition net model

4 实验结果

4.1 样本生成

对通过筛选之后得到的候选聚类，需要生成样本训练聚类识别网络。统计位于真值位置方框中的点云数目，并计算占候选聚类中元素总数的比例，如果比例大于阈值，则认为候选聚类为正样本，即为行人，否则认为候选聚类为负样本。本文在实验中选择的阈值为 0.7。

4.2 数据集介绍

笔者在 KITTI 数据集[21,22]上进行实验，并评估性能指标。KITTI 数据集是知名的自动驾驶数据集，包含双目图像、激光雷达数据以及 IMU 数据，可以对包括识别，跟踪，视觉里程计，光流以及分割等多种任务进行性能指标的评价。KITTI 数据的激光雷达为 Velodyne HDL-64E，拥有 64 根扫描线，最大感知距离为 120 m，水平感知范围是 360°，角分辨率为 0.08°，垂直感知范围为 26.9°，角分辨率大约 0.4°。激光雷达

安装在载具上，高度大约为 1.73 m。目标检测的数据集中含有 7481 张训练数据以及 7518 张测试数据。测试用例分为简单、中等和困难三个等级，其中简单用例为边界框高度不小于 40 个像素且完全无遮挡的目标，中等用例为边界框高度不小于 25 个像素且部分被遮挡的目标，困难用例为边界框高度不小于 25 个像素且大部分被遮挡的目标。

4.3　实验结果

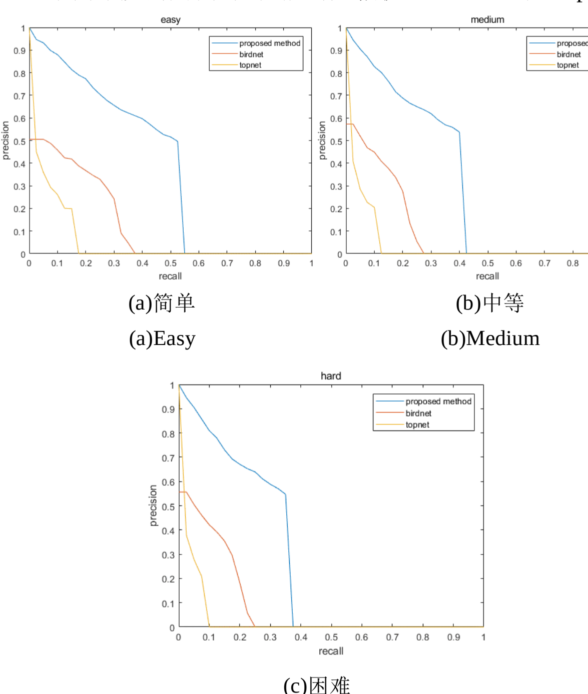
实验使用的计算机核心处理器为 i7-4790K、内存 32 GB、主频 3.1 GHz、显卡为 GeForce GTX TITAN X、显存 12G，实现工具为 Tensorflow r1.4。评价指标为 P-R 曲线和 AP 值。P-R 曲线的横坐标为 recall，纵坐标为 precision，AP 值则是 P-R 曲线下的面积。

$$\text{precision} = \frac{TP}{TP+FN} \tag{7}$$

$$\text{recall} = \frac{TP}{TP+FP} \tag{8}$$

其中: TP 预测结果中的真阳性数，FN 为预测结果中的假阴性数，FP 为预测结果中的假阳性数。

图 5 为本文提出的算法与文献[13,14]结果的 P-R 曲线比较，从左到右分别为简单、中等和困难用例的结果。其中，蓝色曲线为本文提出算法的 P-R 曲线，红色和黄色分别为 BirdNet[13]和 TopNet[14]的结果，可以看出本文提出的算法在处理不同难度的用例时，性能均大幅优于 BirdNet 和 TopNet。



(a)简单　(b)中等

(a)Easy　(b)Medium

(c)困难

(c)Hard

图 5　简单、中等及困难测试用例的 P-R 曲线

Fig. 5　P-R curves belongs to easy, moderate and hard examples

表 1 中为本文提出的算法与 BirdNet 及 TopNet 在 KITTI 数据集的评测结果。可以看到，对于简单及中等的测试用例，本文提出的算法的 AP 值高于 BirdNet 及 TopNet 的结果 20 以上，而对于困难测试用例，本文提出的算法 AP 值高于其他两种方法 10 以上。这充分说明了本文提出的算法在行人检测的任务中优秀性能。

5　结束语

本文提出了结合传统激光雷达运动目标检测和点云分类的行人检测算法，仅利用激光雷达数据，可以直接获取检测目标的深度。算法首先通过地面去除分离地面和目标，接着通过聚类算法获取目标聚类，然后将聚类输入到训练好的深度神经网络中，获得聚类的类别及位置。本文提出的算法能够在 KITTI 数据集上获取良好的结果，性能超过了其他仅利用激光雷达数据的行人检测算法。

表 1　本文算法在 KITTI 数据集上的性能　/%

Table 1　Proposed method performance on KITTI dataset　/%

算法	AP(%)		
	easy	medium	hard
本文算法	**40.19**	**33.37**	**27.90**
BirdNet[13]	14.31	11.80	10.55
TopNet[14]	11.46	10.95	9.09

参考文献：

[1] Jones S M, Viola P. Fast multi-view face detection,TR-20003-96 [R]. Mitsubishi Electric Research Lab 2003.

[2] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 580-587.

[3] Girshick R. Fast R-CNN [C]// Proc of IEEE international Conference on Computer Vision. Washington DC: IEEE Computer Society, 2015: 1440-1448.

[4] Ren Shaoqing, He Kaiming, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. 2015: 91-99.

[5] Zhang Shanshan, Benenson R, Omran M, *et al.* How far are we from solving pedestrian detection? [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2016: 1259-1267.

[6] Liu Fayao, Shen Chunhua, Lin Guosheng. Deep convolutional neural fields for depth estimation from a single image [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2015: 5162-5170.

[7] Garg R, Vijay Kumar B G, Carneiro G, *et al.* Unsupervised cnn for single view depth estimation: Geometry to the rescue [C]//Proc of European Conference on Computer Vision. Springer, Cham, 2016: 740-756.

[8] Asvadi A, Premebida C, Peixoto P, *et al.* 3D lidar-based static and moving obstacle detection in driving environments: #n approach based on voxels and multi-region ground planes [J]. Robotics and Autonomous Systems, 2016, 83: 299-311.

[9] Maturana D, Scherer S. Voxnet: A 3D convolutional neural network for real-time object recognition [C]//Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2015: 922-928.

[10] Qi C R, Su Hao, Mo Kaichun, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2017.

[11] Qi C R, Yi Li, Su Hao, *et al.* PointNet+: deep hierarchical feature learning on point sets in a metric space [C]// Advances in Neural Information Processing Systems. 2017: 5099-5108.

[12] Qi C R, Liu Wei, Wu Chenxia, *et al.* Frustum PointNets for 3D object detection from RGB-D data [EB/OL]. (2018-04-13). https://arxiv.org/pdf/1711.08488.pdf.

[13] Beltran J, Guindel C, Moreno F M, *et al.* BirdNet: a 3D object detection

framework from LiDAR information [J]. arXiv preprint arXiv: 1805. 01195, 2018.

[14] Wirges S, Fischer T, Frias J B, *et al*. Object detection and classification in occupancy grid maps using deep convolutional networks [J]. arXiv preprint arXiv: 1805. 08689, 2018.

[15] Broggi A, Cattani S, Patander M, *et al*. A full-3D voxel-based dynamic obstacle detection for urban scenario using stereo vision [C]//Proc of the 16th International IEEE Conference on Intelligent Transportation Systems. Piscataway, NJ: IEEE Press, 2013: 71-76.

[16] 孙朋朋，赵祥模，徐志刚，等. 基于 3D 激光雷达城市道路边界鲁棒检测算法 [J]. 浙江大学学报:工学版，2018, 52 (3): 504-514. (Sun Pengpeng, Zhao Xiangmo, Xu Zhigang, *et al*. Urban curb robust detection algorithm based on 3D-LIDAR [J]. Journal of Zhejiang University:Engineering Science, 2018, 52 (3): 504-514.)

[17] Oliveira M, Santos V, Sappa A D, *et al*. Scene representations for autonomous driving: an approach based on polygonal primitives [C]// Proc of the 2nd Iberian Robotics Conference. Cham:Springer, 2016: 503-515.

[18] Sappa A D, Herrero R, Dornaika F, *et al*. Road approximation in euclidean and v-disparity space: a comparative study [C]//Proc of International Conference on Computer Aided Systems Theory. Berlin: Springer, 2007: 1105-1112.

[19] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of the ACM, 1981, 24 (6): 381-395.

[20] Ester M, Kriegel H P, Sander J, *et al*. A density-based algorithm for discovering clusters in large spatial databases with noise [C]// Kdd. 1996, 96 (34): 226-231.

[21] Geiger A, Lenz P, Stiller C, *et al*. Vision meets robotics: the KITTI dataset [J]. International Journal of Robotics Research, 2013, 32 (11): 1231-1237.

[22] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving?the kitti vision benchmark suite [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2012: 3354-3361.